THE ANALYST.

VOL. IX. NOVEMBER, 1882. No. 6.

ON AN UNSYMMETRICAL PROBABILITY CURVE.

BY E. L. DE FOREST.

[Continued from page 142.]

Thus it is proved that when any two entire polynomials are multiplied together, the cube of the cubic radius about the centre of forces in the product is equal to the sum of the cubes of the like radii in the two factors. Hence, if any number of such polynomials are multiplied together, the cube of the radius in the final product is equal to the sum of the cubes of the radii in all the factors. The cube of the radius in the k power of a polynomial is k times the cube of the radius for the polynomial itself. These propositions evidently hold true also for polynomials like (5), in which the place of z^0 is not at the first or left hand term. The coefficients in a product are not altered when all the exponents in either factor are increased or diminished by a constant quantity.

Applying the above to the expression for a in (14), we write

$$a = 2\left(\frac{b_2(dx)^2}{b_3(dx)^3}\right) = 2\left(\frac{kb_2(dx)^2}{kb_3(dx)^3}\right), \tag{38}$$

showing that the part within the parentheses may be regarded as the square of the quadratic radius divided by the cube of the cubic radius, either in the first power of the polynomial or in its expansion to the k power, as may be most convenient. The values of a and b may thus be expressed either by means of the coefficients λ in the given polynomial, or by means of the ordinates y to the limiting curve. When the λ's and y's are all positive, and represent probabilities, $kb_2(dx)^2$ is the square of the *quadratic mean error* ε, and $kb_3(dx)^3$ is the cube of what we will call the *cubic mean inequality*, which we denote by ζ. The constants in (14) will then be

$$a = 2\varepsilon^2 \div \zeta^3, \qquad b = \varepsilon^2. \tag{39}$$

The square of the q. m. error is of course positive, but the cube of the c. m. inequality is either + or —, according as the + or — errors preponderate in forming it. If the facility of error is the same on both sides of the arith. mean or centre of gravity, so that + and — errors of equal am't are equally probable, ζ will be zero, and a becomes infinite. If ζ is negative, a is also negative, and the position of the limiting curve is reversed, so that it lies on the left of the origin, or rather, the origin is now at the other side of the expanded polynomial. Since y in (25) is a function of ax, it will have the same value when a and x are both —, as when they are both +.

The notation we adopted in (14) is such that in the expansion of the polynomial (5) to the infinite k power we have

$$(\text{quadratic rad.})^2 = b, \qquad (\text{cubic rad.})^3 = 2b \div a. \tag{40}$$

We will now show that these values of the radii are deducible from the equation (25) of the gamma curve, and thus verify the proposition that the curve is a true limiting form of the expansion of the polynomial, since it possesses properties which are known to characterize that expansion. The distance from the origin to the centre of parallel forces, or centre of gravity of the masses y, will be

$$\frac{1}{dx}\int_0^\infty xy\,dx \div \left(\frac{1}{dx}\int_0^\infty y\,dx\right).$$

The divisor here is unity, so that the distance sought is

$$\frac{1}{dx}\int_0^\infty xy\,dx = \frac{1}{a\Gamma(a^2b)}\int_0^\infty (ax)^{a^2b}\,e^{-ax}d(ax) = \frac{\Gamma(a^2b+1)}{a\Gamma(a^2b)} = ab. \tag{41}$$

This agrees with (12), for by (14) we have

$$ab = 2kb_2^2\,dx \div b_3.$$

The squared quadratic radius of the masses y about the centre of grav. is

$$\frac{1}{dx}\int_0^\infty (x-ab)^2y\,dx \div \left(\frac{1}{dx}\int_0^\infty y\,dx\right),$$

that is, the divisor being unity as before,

$$\frac{1}{dx}\int_0^\infty (x-ab)^2y\,dx = \frac{1}{\Gamma(a^2\,b)}\left\{\frac{1}{a^2}\int_0^\infty (ax)^{a^2b+1}\,e^{-ax}d(ax) - 2b\int_0^\infty (ax)^{a^2b}e^{-ax}d(ax) \right.$$

$$\left. + a^2b^2\int_0^\infty (ax)^{a^2b-1}\,e^{-ax}d(ax)\right\}$$

$$= \frac{1}{\Gamma(a^2b)}\left\{\frac{\Gamma(a^2b+2)}{a^2} - 2b\Gamma(a^2b+1) + a^2b^2\Gamma(a^2b)\right\}$$

$$= b(a^2b+1) - 2a^2b^2 + a^2b^2 = b, \tag{42}$$

a result which agrees with (40). Likewise for the cube of the cubic radius, omitting the divisor unity, we have

$$\frac{1}{dx}\int_0^\infty (x-ab)^3y\,dx = \frac{1}{\Gamma(a^2b)}\left\{\frac{\Gamma(a^2b+3)}{a^3} - \frac{3b\Gamma(a^2b+2)}{a} \begin{matrix} + 3ab^2\Gamma(a^2b+1) \\ - a^3\,b^3\Gamma(a^2b) \end{matrix}\right\}$$

$$= \left(\frac{b}{a}\right)(a^2b+1)(a^2b+2) - 3ab^2(a^2b+1) + 3a^3b^3 - a^3b^3 = \frac{2b}{a}. \qquad (43)$$

This also agrees with (40), so that the curve (25) does exactly represent the form of the series of coefficients in the expansion of the polynomial (5) to an infinitely high power, so far as the quadratic and cubic radii about the centre of forces are concerned.

As shown in my former articles, a binomial $p+q$ or $p+qz$, in which $p+q=1$ and the coefficients p and q are separated by the interval dx, has its centre of gravity at the distance qdx from the first term p, and the sq'd quadratic radius about that centre is

$$(qp^2+pq^2)(dx)^2 = pq(dx)^2. \qquad (44)$$

If the binomial is raised to the mth power, the centre of gravity in the expansion will be at the distance $qmdx$ from the first term, and the squared quadratic radius about that centre is

$$\varepsilon^2 = pqm(dx)^2. \qquad (45)$$

The cube of the cubic radius in the first power is

$$(qp^3-pq^3)(dx)^3 = pq(p-q)(dx)^3, \qquad (46)$$

and in the mth power, as we have here shown, it is m times as great, or

$$\zeta^3 = pqm(p-q)(dx)^3. \qquad (47)$$

Hence, when m becomes infinite, the constants in the limiting curve, according to (39), will be

$$a = \frac{2}{(p-q)dx}, \qquad b = pqm(dx)^2. \qquad (48)$$

We found in (22) that y is a maximum when $x = ab - 1 \div a$, so that the vertex of the gamma curve is at the distance $-1 \div a$ from the centre of gravity, and by (48)

$$-\frac{1}{a} = -\tfrac{1}{2}(p-q)dx. \qquad (49)$$

The agreement between this result and that which I found by different means in ANALYST, Vol. VII, p. 3, shows that the vertex of the curve (25) accurately represents the position of the vertex in the expanded binomial, with reference to the ordinate through the centre of gravity.

It will often be convenient to have the origin of coordinates transferred to the centre of gravity. Putting $x+ab$ in place of x in (25), we have

$$y = \frac{dx}{ab\Gamma(a^2b)}\left(\frac{a^2b}{e}\right)^{a^2b}\left(1+\frac{x}{ab}\right)^{a^2b-1}e^{-ax}. \qquad (50)$$

A known formula for $\Gamma(n)$ is

$$\Gamma(n) = \left(\frac{n}{e}\right)^n\sqrt{\left(\frac{2\pi}{n}\right)}\left(1+\frac{1}{12n}+\frac{1}{288n^2}-\text{\&c.}\right), \qquad (51)$$

and by means of this (50) is reduced to

$$\left.\begin{aligned} K &= 1+\frac{1}{12a^2b}+\frac{1}{288\,(a^2\,b)^2}-\&c., \\ y &= \frac{dx}{K\sqrt{(2\pi b)}}\left(1+\frac{x}{ab}\right)^{a^2b-1} e^{-ax}. \end{aligned}\right\} \qquad (52)$$

This is the equation of the gamma curve referred to the centre of gravity of the masses y as an origin. To illustrate the uses of the curve, we will now employ it in computing the principal terms in the expansion of a binomial to a high power.

Mortality tables show that among persons aged 40, about one per cent may be expected to die within a year, so that the probability of dying within a year is .01, and that of surviving a year is .99. Suppose we wish to find the probabilities that out of 1000 persons aged 40, the number of deaths within a year will be 0, 1, 2, 3 &c. These probabilities are the 1st, 2nd &c. terms in the expansion of the binomial

$$(p+q)^m = (.99+.01)^{1000}. \qquad (53)$$

The common interval between consecutive terms in the expansion being dx, we will take this as the unit of abscissas. By (48) we have

$$a = \frac{2}{.99-.01} = \frac{100}{49}, \quad b = .99\times.01\times1000 = \tfrac{99}{10},$$

and by (52)

$$\log y = \bar{1}.1022148+40.23282\log(1+.04949495x)-.8863153x. \qquad (54)$$

Since the distance $qmdx = 10$ of the centre of gravity of the expanded series from its first term is in this instance a whole number, it follows that one term of the expansion stands exactly at that centre, where $x = 0$, and the whole series of terms will be found by putting x successively equal to

$$\ldots\ldots -2, -1, 0, 1, 2, 3 \ldots\ldots$$

The resulting values of y are given in column (2) of Table I. To show the degree of accuracy attained, the true values of the terms in the expansion have been computed and set in column (1), and the differences (2)—(1) are also given, in units of the fifth decimal place. The computed curve intersects the true one at four points. The agreement between them is pretty close, and would be closer if the exp'nt m were a greater number than it is.

Reverting now to the more general significance of the gamma curve, as representing the expansion of a polynomial, we will inquire what simpler form of curve it approximates to. Let (52) be written

$$y = c\left(1+\frac{x}{ab}\right)^{a^2b-1} e^{-ax},$$

TABLE I.

x	(1)	(2)	(2)-(1)	(3)	(3)-(1)	x	(1)	(2)	(2)-(1)	(3)	(3)-(1)
0	.12574	.12654	80	.12679	105	1	.11431	.11482	51	.12055	624
- 1	.12562	.12635	73	.12055	—507	2	.09516	.09525	9	.10360	844
- 2	.11283	.11311	28	.10360	—923	3	.07305	.07282	—23	.08048	743
- 3	.08999	.08966	—33	.08048	—951	4	.05202	.05167	—35	.05651	449
- 4	.06274	.06203	—71	.05651	—623	5	.03454	.03422	—32	.03587	133
- 5	.03746	.03680	—66	.03587	—159	6	.02148	.02127	—21	.02058	— 90
- 6	01861	.01834	—27	.02058	197	7	.01256	.01247	— 9	.01067	—189
- 7	.00739	.00748	9	.01067	328	8	.00693	.00692	— 1	.00500	—193
- 8	00220	.00242	22	.00500	280	9	00362	.00365	3	.00212	—150
- 9	.00044	.00060	16	.00212	168	10	.00179	.00184	5	.00081	— 98
-10	.00004	.00011	7	.00081	77	11	.00084	.00089	5	.00028	— 56
-11		.00002		.00028		12	.00038	.00041	3	.00009	— 29
-12				.00009		13	.00016	.00018	2	.00002	— 14
-13				.00002		14	.00007	.00008	1	.00001	— 6
-14				.00001		15	.00003	.00003	0		— 3
						16	.00001	.00001	0		— 1
						17		.00001	1		

$$\therefore \log'\left(\frac{y}{c}\right) = (a^2b-1)\log'\left(1+\frac{x}{ab}\right) - ax$$

$$= (a^2b-1)\left\{\frac{x}{ab}-\frac{1}{2}\left(\frac{x}{ab}\right)^2+\frac{1}{3}\left(\frac{x}{ab}\right)^3-\&c.\right\}-ax$$

$$= -\frac{x^2}{2b}+\left(\frac{x^2}{3b}-\frac{1}{1}\right)\frac{x}{ab}-\left(\frac{x^2}{4b}-\frac{1}{2}\right)\left(\frac{x}{ab}\right)^2+\left(\frac{x^2}{5b}-\frac{1}{3}\right)\left(\frac{x}{ab}\right)^3\ldots \quad (55)$$

Since b is a finite area, $x^2\div b$ is in general a finite number. By (14) we have

$$\frac{x}{ab} = \left(\frac{b_3}{b_2}\right)\frac{xdx}{2b}, \quad (56)$$

which is in general an infinitesimal. Neglecting all terms in which $x\div ab$ is a factor, (55) is reduced to

$$\log'\left(\frac{y}{c}\right) = -\frac{x^2}{2b}, \qquad \therefore y = ce^{-x^2\div 2b}.$$

Restoring the value of c from (52), noticing that when k is really infinite we have

$$a^2b = 4kb_2^3\div b_3^2 = \infty, \qquad \therefore K = 1,$$

we get finally

$$y = \frac{dx}{\sqrt{(2\pi b)}}e^{-x^2\div 2b}, \quad (57)$$

the equation of a common probability curve like (1). This is the same result we would have obtained in the first place, if we had negl'd d^2y in (10).

If instead of retaining only dy and d^2y, we should also retain d^3y, the resulting equation, if we could integrate it, would doubtless give a limiting curve of still more general form, of which the gamma curve is but a particular case. Under this view, the probability curve (57) and the gamma curve (52) are only first and second approximations to the actual form of an expansion to a high power.

It must be observed that since dx represents the common interval between consecutive coefficients y in the expanded series, the abscissas x corresponding to ordinates at and near the origin will have the values

$$\ldots\ldots -dx,\ 0,\ dx,\ 2dx,\ \ldots\ldots,$$

so that $x^2 \div b$ and $x \div ab$ in (55) will there be of the same order of magnitude, and the latter cannot be neglected in comparison with the former. The curve is thus rendered unsymmetrical in the immediate vicinity of the origin, and the maximum or vertex is thrown a very little to one side of the centre of gravity. But if the given polynomial (5) is such as to make $b_3 = 0$, then (56) makes $x \div ab$ absolutely null, and the terms in which it is a factor disappear altogether from (55), leaving (57) as the exact result, and showing that the probability curve is a special case of the gamma curve, occurr'g when $1 \div a = 0$. This case arises at the moment when the gamma curve passes from the direct to the reversed position, as noticed in connection with (39). The X axis then becomes an asymptote to the curve, not on one side only, but on both sides. Again, if b_3, though not absolutely zero, is quite small in comparison with b_2, making ζ^3 quite small in comparison with ε^2, a thing which often occurs, then a is a very large number, and terms containing $x \div ab$ may be dropped from (55) as before. In other words, if the asymmetry, as measured by the c. m. inequality ζ, is small, the gamma curve does not differ materially from the common probability curve, and the latter may be preferred to it for practical use, as being more simple.

But when ζ is so large that a is a small number, the probability curve will not represent the true form of the expansion of a polynomial to a finite power with sufficient accuracy. Take for instance that of the binomial (53), for which, with $dx = 1$ as before, and $b = 9.9$, we get by (57)

$$\log y = \bar{1}.1030924 - .02193407x^2. \qquad (58)$$

The terms of the series computed by this are entered in column (3) of Table I. Their sum is unity as it should be. The differences (3) — (1) between the computed values and the true ones are also shown. On an average they are more than 12 times as great as the differences (2)–(1) afforded by the gamma curve.

That the expansion of a polynomial to a high power, taken as a whole, tends to become more and more symmetrical in form the higher the power

is, may be inferred from the properties of the quadratic and cubic radii, without regard to any precise analytical expression for the limiting curve. While the whole length of the expanded series increases in proportion to the exponent k of the power, the quadratic radius of the coefficients, about their centre of forces, increases only as the square root of k, while the cubic radius increases still more slowly, being proportional to the cube root of k, as seen in connection with (37).

Suppose that the coefficients λ in the polynomial (5) represent the probabilities of the occurrence of the various possible true errors of an observed quantity, these errors being multiples of the unit of measure $\varDelta x$, which may be taken as small as we please, while m is a whole number so large that the greatest error will not exceed $\pm\, m\varDelta x$. In any term $\lambda_i z^i$ the coefficient λ_i is the probability that, in a single observation, the error which occurs will be $x = i\varDelta x$. The centre of gravity of the coeffic's, regarded as the masses of material p'ts ranged along the imponderable axis of X, may or may not coincide with the place of λ_0, but wherever it is, its abscissa x_1, or lever arm about the place of λ_0, is the arithmetical mean of all the possible true errors of a single observation, each error being taken with a weight proportional to the probability of its occurrence. Let ε and ζ denote the quadratic and cubic radii of the masses λ, about the centre of gravity. These are the same as the q. m. error and c. m. inequality of a single observation, if what we call errors are not necessarily true errors, but only deviations from the centre of gravity or *ultimate arith. mean.* (ANALYST, VIII, p. 141.)

If k such observations are taken, the possible true errors of their sum will be the exponents, and their probabilities will be the coefficients l, in the polynomial (6), which is the expansion of (5) to the k power. The centre of gravity of all the coefficients l will be approximately the place of the maximum coefficient, and its abscissa, or lever arm about the place of l_0, will be kx_1. This lever arm is the arithmetic mean of all the possible true errors in the sum of k observations, the errors being weighted for probability of occurrence. The quadratic and cubic radii for the masses l, about the centre of gravity, will be

$$E = \varepsilon\sqrt{k}, \qquad Z = \zeta\sqrt[3]{k}, \tag{59}$$

and these are the q. m. error and c. m. inequality of the sum of k observations.

The probability that the true error of the sum of k observations will be x, is the same as the probability that the true error of their arith. mean will be $x\div k$. If we suppose the coefficients or masses l to be set closer together, so that the common interval between them is reduced from $\varDelta x$ to $\varDelta x\div k$, their distribution along the X axis will represent the law of facility of error in the mean of k observations. The limits of possible error, which were $\pm\, km\varDelta x$ for the sum, will be reduced to $\pm\, m\varDelta x$ for the mean, being the same

as for a single observation. The centre of g. of all the masses l will now, as before, be the approximate place of the maximum, and its abscissa, or lever arm about the place of l_0, will be reduced from kx_1 to x_1, showing that if there is any true error in the ultimate mean, or arith. mean of all the possible values, weighted for probabi'ty of occurrence, it is the same for the arith. mean of k observations, as it is for a single observation.

Since the distance of each of the masses l from the centre of gravity is k times less for the arith. mean than it was for the sum, $1 \div k$ becomes a common coefficient of all the distances which go to make up the quadratic and cubic radii, which are consequently k times less for the mean than they were for the sum. Hence by (59), the q. m. error and c. m. inequality for the arith. mean of k observations will be

$$\varepsilon_0 = \varepsilon \div \sqrt{k}, \qquad \zeta_0 = \zeta \div k^{2/3}. \tag{60}$$

When k is increased, the q. m. error of the mean diminishes, being inv'sely as $\sqrt{k}$, while the c. m. inequality diminishes more rapidly, being inversely as $\sqrt[3]{k^2}$. If we take 64 times as many observations, the q. m. error of the mean result will be one eighth as large as before, but the c. m. inequality of the possible errors of the mean will be only one sixteenth as large as before. This goes to show that as k increases, the curve of facility of error in the mean, taken as a whole, becomes more and more symmetrical on either side of the centre of gravity.

It may be remarked here, by the way, that a relation holds for the c. m. inequality, very similar to that which holds for the q. m. error, in a quantity X which is connected with other quantities x_1, x_2 &c. thus,

$$X = a_1x_1 + a_2x_2 + a_3x_3 + \&c., \tag{61}$$

where a_1, a_2 &c. may be essentially either + or —. The errors of x_1, x_2 &c. are supposed to be independent, that is, the error of one has no influence on the error of another. If ε_1, ε_2 &c. denote the q. m. errors of x_1, x_2 &c., and ε denotes the q. m. error of X, then as is well known

$$\varepsilon^2 = (a_1\varepsilon_1)^2 + (a_2\varepsilon_2)^2 + (a_3\varepsilon_3)^2 + \&c. \tag{62}$$

(See the method of proof which I gave in ANALYST, VIII, p. 139.) In like manner, it is easily seen that if the errors to which x_1, x_2 &c. are liable are of such nature that + and — errors of equal amount are not equally probable, then denoting the c. m. inequalities of x_1, x_2 &c. by ζ_1, ζ_2 &c., and that of X by ζ, we shall have by virtue of (37)

$$\zeta^3 = (a_1\zeta_1)^3 + (a_2\zeta_2)^3 + (a_3\zeta_3)^3 + \&c. \tag{63}$$

The a_1, a_2 &c. are merely coefficients, so that any actual error of a_1x_1, for instance, is a_1 times the actual error of x_1. Then the c. m. inequality of a_1x_1 is $a_1\zeta_1$, that of a_2x_2 is $a_2\zeta_2$, and so on. The c. m. inequality of $-x_2$ being $-\zeta_2$, that of $(x_1 \pm x_2)$ is the cube root of $(\zeta_1^3 \pm \zeta_2^3)$.

[To be continued.]